BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
SECOND SEMESTER – 2016-17

Course No.: ECON F342 **Course Title: APPLIED ECONOMETRICS**
Date: 12 MAY 2015 **Max. Marks: 40**

COMPREHENSIVE EXAMINATION

Instructions: Open Book Examination. Do all the following problems. Please write legibly and organize your notation and arguments clearly. Write assumptions if any clearly. Start answering each question on a fresh page. Attempt all parts of the question at one place.

1). We are interested in explaining a worker's wage in terms of the number of years of education (educ) and years of experience (exper) using the following model:

$$log\,(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + u,$$

where we assume that u satis.es the classical assumptions of OLS and is homoskedastic. The estimated parameters by OLS for a sample of n = 935 observations are displayed in the first column of Table

TABLE: OLS Estimates – Dependent variable : log (wage)

Variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
educ	0.07778	0.05316	0.07815	0.07192	0.071984	0.05571
	(0.00669)	(0.02085)	(0.00653)	(0.00666)	(0.00677)	(0.00600)
exper	0.01977	0.00038	0.01829	0.01791	0.01678	
	(0.00330)	(0.01066)	(0.00330)	(0.00327)	(0.01389)	
$educ * married$		0.02813				
		(0.02194)				
$exper * married$		0.01952				
		(0.01120)				
married		-0.38069	0.20926	0.18886	0.18873	0.21311
		(0.36818)	(0.04272)	(0.04764)	(0.04763)	(0.04709)
black				-0.24167	-0.24128	-0.22500
				(0.08391)	(0.08417)	(0.08212)
$married * black$				0.03599	0.03543	0.01071
				(0.09387)	(0.09404)	(0.09224)
$exper^2$					4.86e-05	
					(0.00058)	
const	5.50271	5.85694	5.32796	5.46166	5.46653	5.86609
	(0.11427)	(0.34889)	(0.11574)	(0.12017)	(0.12914)	(0.09445)
Observations	935	935	935	935	935	935
R^2	0.13086	0.15705	0.15420	0.18131	0.18132	0.15417

Several extensions of this model were considered to address the effects of being married (with the binary variable married) and/or being black (with binary black) or possible nonlinearity on the effects of years of experience. Using the appropriate output from Table answer the following questions:

i Test whether the wage regressions for married workers and unmarried workers are the same. Write down the estimated regression for each group and interpret the coefficient of educ.

ii Based on a statistical test, do the effects of education and experience depend on the civil status?

iii What we conclude about the possible nonlinearity of the relationship of log (wage) with respect to the years of experience? Can you conclude that years of experience has no significant effect on log (wage) in Model 5? Make two statistical tests to answer these questions. **(6.0)**

2). Romer (1993) proposes theoretical models of inflation that imply that more “open” countries should have lower inflation rates. His empirical analysis explains average annual inflation rates (since 1973) in terms of the average share of imports in gross domestic product since 1973 (which is a measure of openess). In addition to estimate the key equation by OLS, he uses instrumental variables. While Romer does not specify both equations in a simultaneous system, he has in mind a two-equation system:

$$in = \beta_0 + \beta_1 open + \beta_1 \log(pcinc) + u_1 \text{ ....(1)}$$

$$open = \gamma_0 + \gamma_1 in + \gamma_2 \log(pcinc) + \gamma_3 \log(land) + u_2 \text{....................(2)}$$

Where pcinc is 1980 per capita income, in U.S. dollars (exogenous), and land is the land area of the country, in square miles (exogenous).

The first equation is the one of interest; with the hypothesis that $\beta_1 < 0$ (more open economies have lower inflation rates).

The second equation reflects the fact that the degree of openess might depend on the average inflation rate, as well as other factors.

We first estimate the reduced form regression:

$$\underset{(7.38)}{open = 117.08} + \underset{(0.36)}{0.546} \log(pcinc) - \underset{(9.34)}{7.57} \log(land)$$

And finally estimate the first equation using log(land) as an IV for open

$$in = \underset{(1.75)}{26.90} - \underset{(2.34)}{0.337} open + \underset{(0.19)}{0.376} \log(pcinc).$$

i Is the first equation identified? Is the second equation identified?

ii Interpret the coefficients from the reduced form regression.

iii Interpret the coefficients from the last regression.

iv Test the null hypothesis that more openess results in less inflation.

(8.0)

3). Consider the following panel data (small) set of observations for Y_{it}.

		t		
		1	2	3
i	1	15	18	15
	2	11	17	11
	3	13	19	10

i What is the entity demeaned value of Y for entity 3 in time 3?

ii What is the time demeaned value of Y for entity 2 in time 3?

iii What is the entity and time demeaned value of Y for entity 1 in time 3?

(4.0)

4).

A Suppose you have a model of the form $Y_i = \beta X_i + u_i$,

$$X_i = \pi Z_i + v_i \text{ ,}$$

where (Y_i , X_i , Z_i) are i.i.d.; $Cov(u_i , v_i) \neq 0$; $\pi \neq 0$; and, with probability one, $E[v_i | Z_i] = 0$. In addition, assume that $E[Y_i^4] < \infty$, $E[Z_i^4] < \infty$, and $E[X_i^4] < \infty$. You are not sure, however, whether the condition Cov (Zi , ui) = 0 holds or not. Compute the J-statistic. Can you test the exogeneity assumption using the J-test in this case? Briefly explain.

B The following model is estimated using a balanced panel of five firms over 20 years: $I_{it} = \beta_1 F_{it} + \beta_2 C_{it} + u_{it}$. where the regressors are market value (F) and capital (C) and the dependent variable is investment (I). Suppose that the true error structure of the model is "$u_{it} = \alpha_i + \eta_{it}$, where α is uncorrelated with the regressors.

i If the model is estimated as a fixed effects model, what will be the statistical properties, in terms of efficiency and consistency, of the estimates?

ii The estimates for pooled OLS, fixed effects (using dummies) and random effects models are given in the table below. Use the statistics shown to decide whether the data support a fixed effects or random effects specification. Carefully explain your reasoning.

Dependent Variable is Investment

Estimation	Constant	Market Value	Capital
(a) OLS	-48.030 (-2.236)	0.10509 (9.236)	0.30537 (7.019)
(b) Fixed Effects	-	0.10598 (6.669)	0.34666 (14.348)
(c) Random Effects	-61.575 (-0.775)	0.10549 (6.859)	0.34641 (14.350)

(t-ratios are shown in brackets)

Breush-Pagan LM test for random effects (1 df): 453.82

Hausman test of fixed vs random effects (2 df): 1.27

(6.0)

5). The following model is a system of simultaneous equations to study whether the openness of the economy (open) leads to lower inflation rates (inf),

$$\begin{aligned} inf &= \delta_{10} + \gamma_{12} open + \delta_{11} \log(pcinc) + u_1 \\ open &= \delta_{20} + \gamma_{21} inf + \delta_{21} \log(pcinc) + \delta_{22} \log(land) + u_2. \end{aligned}$$

We assume that (the logarithms of) pcinc (per capita income) and land (land for farming) are exogenous in the whole exercise. The following estimations have been obtained by OLS and 2SLS .

i Discuss the possible identification of each equation of the system, the weakness of the available instruments and perform the correspondent hypothesis tests whenever is possible.

ii Explain how you would perform a test of the exogeneity of the instruments used in the two-stage estimation for a equation and whether it is possible to apply it for the equations of the given system.

iii Test whether the effect of open over inf is lower than -0:2: If open were not a determinant of inf, (but inf is a determinant of open), explain the properties of the estimates of Output 1.

(8.0)

Output 1: OLS estimation using the 114 observations 1–114
Dependent variable: inf

Variable	Coefficient	Standard Dev.	t statistic	p-value
const	25,1040	15,2052	1,6510	0,1016
open	-0,215070	0,0946289	-2,2728	0,0250
lpcinc	0,0175673	1,97527	0,0089	0,9929

Mean of dependent variable	17,2640
Std. dev. of dependent variable	23,9973
Residual sum of squares	62127,5
Residual standard deviation ($\hat{\sigma}$)	23,6581
R^2	0,0452708
$\bar{R}^2$ corrected	0,0280685
$F(2, 111)$	2,63167
p-value for $F()$	0,0764453

Output 2: OLS estimation using the 114 observations 1–114
Dependent variable: open

Variable	Coefficient	Standard Dev.	t statistic	p-value
const	116,226	15,8808	7,3187	0,0000
inf	-0,0680353	0,0715556	-0,9508	0,3438
lpcinc	0,559501	1,49395	0,3745	0,7087
lland	-7,3933	0,834814	-8,8563	0,0000

Mean of dependent variable	37,0789
Std. dev. of dependent variable	23,7535
Residual sum of squares	34865,3
Residual standard deviation ($\hat{\sigma}$)	17,8033
R^2	0,453162
$\bar{R}^2$ corrected	0,438249
$F(3, 110)$	30,3855
p-value for $F()$	$< 0,00001$

Output 3: OLS estimation using the 114 observations 1–114
Dependent variable: inf

Variable	Coefficient	Standard Dev.	t statistic	p-value
const	-12,615	21,0313	-0,5998	0,5498
lpcinc	0,191394	1,98158	0,0966	0,9232
lland	2,55380	1,08049	2,3635	0,0198

Mean of dependent variable	17,2640
Std. dev. of dependent variable	23,9973
Residual sum of squares	61903,2
Residual standard deviation ($\hat{\sigma}$)	23,6154
R^2	0,0487174
$\bar{R}^2$ corrected	0,0315772
$F(2, 111)$	2,84229
p-value for $F()$	0,0625432

Output 4: OLS estimation using the 114 observations 1–114
Dependent variable: open

Variable	Coefficient	Standard dev.	t statistic	p-value
const	117,085	15,8483	7,3878	0,0000
lpcinc	0,546479	1,49324	0,3660	0,7151
lland	-7,5671	0,814216	-9,2937	0,0000

Mean of dependent variable	37,0789
Std. dev. of dependent variable	23,7535
Residual sum of squares	35151,8
Residual standard deviation ($\hat{\sigma}$)	17,7956
R^2	0,448668
$\bar{R}^2$ corrected	0,438734
$F(2, 111)$	45,1654
p-value for $F()$	$<$0,00001

Output 5: 2SLS estimation using the 114 observations 1–114
Dependent variable: inf
Instruments: lland

Variable	Coefficient	Standard dev.	t statistic	p-value
const	26,8993	15,4012	1,7466	0,0807
open	-0,337487	0,144121	-2,3417	0,0192
lpcinc	0,375823	2,01508	0,1865	0,8520

Mean of dependent variable	17,2640
Std. dev. of dependent variable	23,9973
Residual sum of squares	63064,2
Residual standard deviation ($\hat{\sigma}$)	23,8358
$F(2, 111)$	2,62498
p-value for $F()$	0,0769352

Hausman Test –
Null hypothesis: OLS estimates are consistent
Asymptotic test statistic: $\chi^2_1 = 1{,}35333$
with p-value = 0,244697

6). Taylor Swift wrote an op-ed article in the The Wall Street Journal speaking out against pirating (i.e., illegally downloading) music and made news again recently because of her decision to remove her music from Spotify, an online streaming music website, The New York Times reported. Responding to critics, Daniel Ek, Spotify's CEO, states Spotify is ". . . a platform that protects them [music artists] from piracy and pays them for their amazing work." Musical piracy is an increasingly popular topic of debate with policymakers having to decide whether or not to change laws in order to adapt to rapidly changing technology. Further, the music industry puts more pressure on policymakers in order to protect their 15 billion dollar a year industry. As a consequence, economists have examined the impact of music piracy on music sales. Essentially, the following studies want to estimate an equation along the lines of

$$\textbf{sales} = \beta_0 + \beta_1\textbf{piracy} + \textbf{X}\beta + \textbf{u}$$

where sales is quantity of album sales, piracy is a variable representing the amount of illegally downloaded music, X is a matrix of explanatory (control) variables , and u is the idiosyncratic error term.

A Anderson and Frenz (2010) use Canadian survey data and an instrumental variable approach to estimate a causal relationship of Equation (1). The paper finds illegal file sharing has a negligible effect (i.e., $\beta_1 \approx 0$) on music sales arguing such activities create a range of new business opportunities. Findings from the paper have been used as expert evidence in two British landmark court-cases dealing with peer-to-peer (P2P) file-sharing

i. Assuming the data used in the paper are sound. Why are the authors unable to use ordinary least squares (OLS) to obtain causal estimates? Explain

ii. The paper instruments for piracy by using internet skills (skills) whereby a respondent reports their respective level of internet expertise (sophistication). The first-stage regression is

$$\underset{(\text{robust } \hat{se})}{\widehat{piracy}} = \underset{(0.339)}{0.037} + \underset{(0.142)}{0.790}\,skills - \underset{(0.010)}{0.002}\,price - \underset{(0.146)}{0.030}\,student + \underset{(0.080)}{0.036}\,female - \underset{(0.077)}{0.249}\,region$$

What are the two important criteria for a valid ("good") instrumental variable (IV)? Can the two criteria be tested? Does the authors' IV meet these two criteria? Explain

B Another researchers Adermon and Liang (2014) use a difference-in-differences (DID) approach, with Norway and Finland as control groups, exploiting the implementation of copyright protection reform in Sweden (i.e., treatment group) in April 2009 that suddenly increased the risk of being caught and punished for illegal file sharing. The paper's DID model is estimated with a regression to allow for controls:

$$\underset{(\text{robust } \hat{se})}{\ln(\widehat{sales})} = \underset{(0.0886)}{0.0942} + \underset{(0.0861)}{0.455}\,Sweden + \underset{(0.0960)}{0.254}\,Post2009 + \underset{(0.0722)}{0.364}\,reform + \boldsymbol{X}\hat{\beta} \qquad (2)$$

where the log of the continuous variable sales is the dependent variable, Sweden is the treated group dummy, Post2009 is a dummy for the period after the April 2009 policy change, reform is the dummy variable indicating treatment (i.e., Sweden_Post2009), and X is a matrix of explanatory (control) variables (Note: $\hat{\beta}$ are the respective estimates (not shown.).

i What is the key assumption underlying the validity of a DID estimate to be considered causal? Can it be tested? Explain. What is the estimated effect of Sweden's copyright protection reform? Use the coefficient and comment on statistical significance at the 95% confidence level.

ii The paper uses robust standard errors. What are the authors concerned about; i.e., why are robust standard errors used? What would the consequences be if the problem the authors are worried about is not addressed? Explain.

(8.0)